COMMUNICATION

ArrayOU: A Web Application for Microarray Data Analysis and Visualization

Kaiyu Shen,[1,2] *Sarah E. Wyatt,*[1] *and Vijayanand Nadella*[2,*]

[1]*Department of Environmental and Plant Biology and* [2]*Ohio University Genomics Facility, Ohio University, Athens, Ohio 45701, USA*

A web-based microarray data analysis tool, ArrayOU (freely available at www.bioinformatics.plantbio.ohiou. edu.), has been developed at the Ohio University Genomics Facility for the research and education community to analyze Agilent microarray data. Agilent's microarray pipeline has gained in popularity as a result of its ease of use and low cost of customized arrays. The current version of the ArrayOU pipeline allows users to visualize, analyze, and annotate microarray data from commercially available and customized Agilent expression arrays and is extendable for further implementations.

KEY WORDS: DNA microarrays, software, pipeline

INTRODUCTION

Oligonucleotide arrays have become popular to study gene expression in a wide range of organisms. Among the most popular platforms, Agilent technology microarray uses 60-mer oligonucleotide probes,[1] and the targets can be labeled with single or dual-color dyes. In addition to the commercially available arrays for model organisms, custom arrays can be printed through Agilent printing technology for no extra cost, increasing the flexibility of the system.

However, microarray data analysis poses challenges for many biologists. Although there is commercial software available for microarray data analysis, these software choices are often limited in their options or parameter settings for in-depth data analysis. Besides the limitations, the licenses for some of the commercial software have to be renewed annually, making it cost-prohibitive to some researchers. Bioconductor,[2] on the other hand, is a collection of R packages written specifically for microarray data analysis and annotation, is available for no cost, and provides excellent functionalities for researchers to analyze their data. However, use of the Bioconductor packages requires the ability to understand and write R scripts. ArrayOU addresses these issues of functionality and ease of operation to help researchers with analysis of microarray data. The software was designed to provide a stepwise pipeline and freedom for users to determine parameter settings in an interactive way.

MATERIALS AND METHODS

The ArrayOU pipeline integrates a collection of Bioconductor packages and BioPerl modules. The website adopts hypertext preprocessor (PHP) and hypertext markup language (HTML) as its main framework and uses Asynchronous JavaScript XML (AJAX) technology to make the communication between the server and client more dynamic. R scripts and BioPerl modules were also installed and compiled in the server. MySQL is used to communicate with the database and record the entries and the results of the registered users.

RESULTS AND DISCUSSION

ArrayOU provides a stepwise pipeline for microarray data analysis (Fig. 1). Raw data (single-color or dual-color arrays; text-formatted files) are uploaded (multiple uploads are supported) into the pipeline and processed. The platform adopts a Linear Models for Microarray Data (LIMMA)[3,4] package to perform background correction and normalization.[5] A corresponding quality control (QC) report is generated by the arrayQuality package (http://www.bioconductor.org/packages/release/bioc/html/arrayQuality.html), including a series of dual-color intensity (M):average intensity (A) ratio (MA) plots for each dual-color array (Fig. 2A) and single-channel intensity plots or box plots for overall expression levels for all single-color arrays (Fig. 2B). Users can adjust the parameter settings based on the QC reports. Background correction is performed to eliminate noise, and expression values are normalized within arrays (specifically for dual-color arrays) or between arrays to bring the data to the same scale. The genefilter (http://www.bioconductor.org/packages/release/bioc/html/genefilter.html) package is used to select the
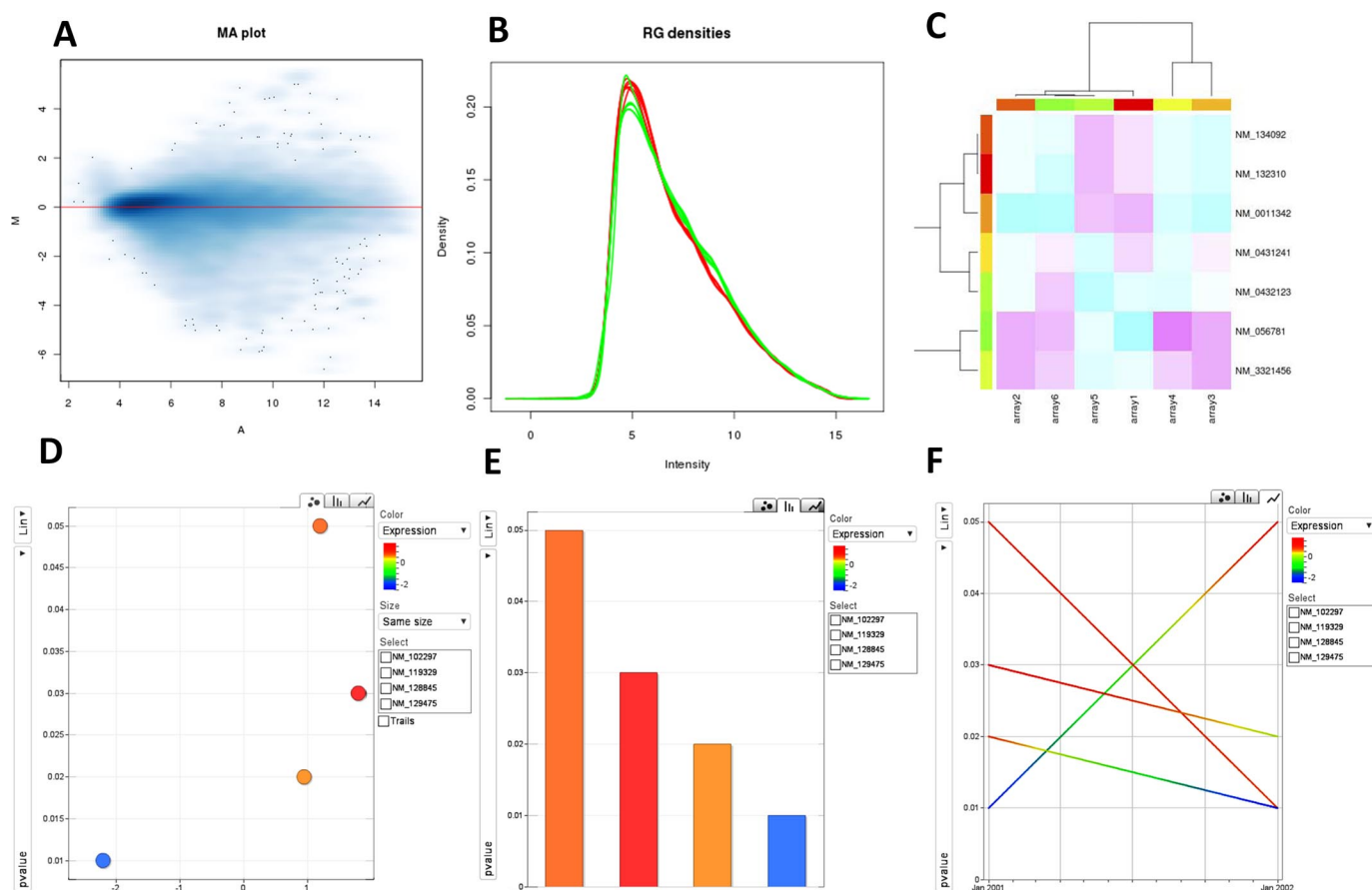
FIGURE 1

The ArrayOU Pipeline. Single-direction arrows show the flow of data. Double-direction arrows represent custom settings. Rectangular boxes indicate the major steps of analysis; diamond boxes show crucial results in the pipeline; and ellipse boxes represent the plots and analysis reports.
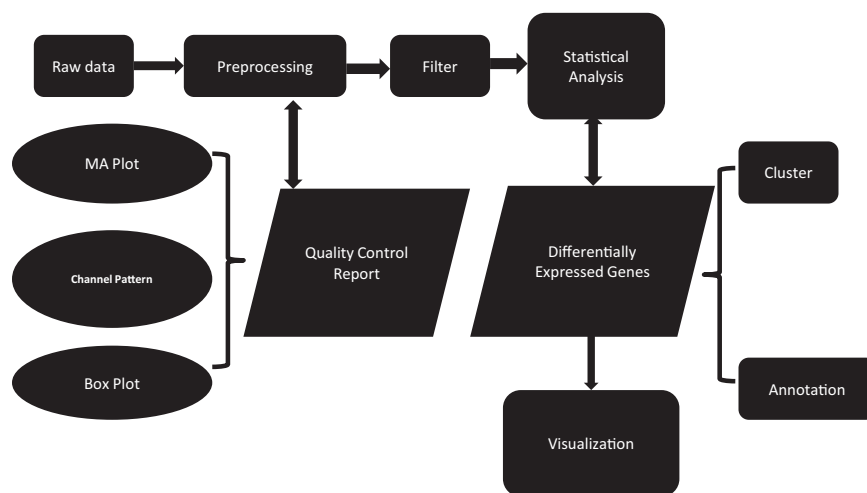
genes within a designated quantile in any number of samples, which ensures the statistical power, and the list of genes generated will be used for further analysis.

Based on the nature of the experimental design, most microarray data fall into two categories: factorial and time course.[6] Different statistical methods can be applied as appropriate for the experimental design. ArrayOU provides options for the empirical Bayesian[7] model in a LIMMA package and the nonparametrical model in the Rank-Product[8] package to perform statistical analyses on gene-expression values. Researchers may choose from a variety of statistical tests, available from a drop-down menu, to select

FIGURE 2

The dynamic visualization of gene-expression profiles. (*A*) The MA plot for dual-color array. (*B*) Single-channel density plot. (*C*) Heatmap plot of cluster genes. (*D–F*) Dot plot, bar plot, and line plot for expression profiles.

for statistically significant, differentially expressed genes by assigning the *P* value and log-fold change cutoff.

The genes of interest can be annotated using public-available databases and clustered based on expression profiles. ArrayOU now provides annotation for *Homo sapiens*, *Arabidopsis thaliana*, and *Mus musculus* with additional species annotations to be added. Alternatively, genes of interest can be clustered based on expression patterns (Fig. 2C). Through BioPerl modules, ArrayOU allows users to select the promoter regions of the genes of interest, compare gene sequence similarities, and retrieve information from PubMed and other databases. ArrayOU also provides a dynamic visualization on the gene-expression profiles, which allows the user to input a list of gene names and visualize their expression levels across all experimental sets. This dynamic visualization adopts Google Visualization [application programming interface] API Reference (http://code.google.com/intl/en/apis/chart/interactive/docs/reference.html) and provides a line graph, bar graph, and dot graph on the genes that the user selects (Fig. 2D–F). This dynamic visualization provides an easier and more straightforward way to identify gene-expression changes between time-points in a time-course manner or a comparison between control and treatment in a factorial experiment.

ArrayOU adopts multiple options and renders the users with great flexibility in analyzing the microarray data. This pipeline has been applied to analyze real data sets, which proved to be helpful in finding and visualizing differentially expressed genes.

REFERENCES

1. Leiske DL, Karimpour-Fard A, Hume PS, Fairbanks BD, Gill RT. A comparison of alternative 60-mer probe designs in an in-situ synthesized oligonucleotide microarray. *BMC Genomics* 2006;7:72.
2. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;5:R80.
3. Ritchie ME, Silver J, Oshlack A, et al. A comparison of background correction methods for two-colour microarrays. *Bioinformatics* 2007;23:2700–2707.
4. Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 2005;21:2067–2075.
5. Smyth GK, Speed T. Normalization of cDNA microarray data. *Methods* 2003;31:265–273.
6. Glonek GF, Solomon PJ. Factorial and time course designs for cDNA microarray experiments. *Biostatistics* 2004;5:89–111.
7. Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004;3:Article3.
8. Breitling R, Armengaud P, Amtmann A, Herzyk P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett* 2004;573:83–92.